

Konfirmatorische Analysen der Reliabilität von Wichtigkeitseinstufungen beruflicher Merkmale

Faulbaum, Frank

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Faulbaum, F. (1983). Konfirmatorische Analysen der Reliabilität von Wichtigkeitseinstufungen beruflicher Merkmale. *ZUMA Nachrichten*, 7(13), 22-44. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-210479>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

KONFIRMATORISCHE ANALYSEN DER RELIABILITÄT VON WICHTIGKEITSEINSTUFUNGEN BERUFLICHER MERKMALE

1. Problemstellung

Der vorliegende Aufsatz basiert auf den Ergebnissen der Analyse einer Teilfragestellung der Methodenstudie "Skalenvergleich", die als Zusatzstudie zum ALLBUS 1982 konzipiert wurde und deren zentrales Thema die Vergleichbarkeit von Einstellungsskalen im deutschen ALLBUS und im amerikanischen GENERAL SOCIAL SURVEY (GSS) ist.* In Abstimmung mit dem National Opinion Research Center (NORC) in Chicago, das den GSS in den Vereinigten Staaten betreut, wurde die Frage nach der Vergleichbarkeit von Einstellungsmessungen exemplarisch eingegrenzt auf die Itembatterie zur subjektiven Wichtigkeit von Merkmalen der beruflichen Arbeit (ALLBUS 1982, Frage 6A; GSS 1982, Frage 84Y). Dabei sollten die folgenden 13 Merkmale der beruflichen Arbeit u. a. auf einer 7-stufigen Kategorienskala hinsichtlich der subjektiven Wichtigkeit für den Beruf und die berufliche Arbeit eingestuft werden:

Sichere Berufsstellung

Hohes Einkommen

Gute Aufstiegsmöglichkeiten

Ein Beruf, der anerkannt und geachtet wird

Ein Beruf, der einem viel Freizeit läßt

Interessante Tätigkeit

Eine Tätigkeit, bei der man selbständig arbeiten kann

Aufgaben, die viel Verantwortungsbewußtsein erfordern

Viel Kontakt zu anderen Menschen

Ein Beruf, bei dem man anderen helfen kann

Ein Beruf, der für die Gesellschaft nützlich ist

Gibt mir das Gefühl, etwas Sinnvolles zu tun

Sichere und gesunde Arbeitsbedingungen

*Die Methodenstudie wurde mit Mitteln der Deutschen Forschungsgemeinschaft finanziert. Antragsteller waren Professor M. Rainer Lepsius, Universität Heidelberg, gemeinsam mit Professor Erwin K. Scheuch, Universität Köln, und Professor Rolf Ziegler, Universität München.

Während im amerikanischen Teil der Untersuchung nur die numerische Version der 7-stufigen Wichtigkeitsskala zur Anwendung kam, wurde im deutschen Teil zusätzlich noch eine verbale Version der Skala verwendet, wobei den Wichtigkeitsstufen 1 - 7 die verbalen Abstufungen

"nicht wichtig" (1), "wenig wichtig" (2), "etwas wichtig" (3), "einigermaßen wichtig" (4), "ziemlich wichtig" (5), "überwiegend wichtig" (6) und "sehr wichtig" (7)

entsprachen. Die Adverbien "nicht", "etwas" etc. hatten sich in einer Untersuchung von ROHRMANN (1978) als relativ gleichabständig herausgestellt.

Im Zusammenhang mit den beiden Skalenversionen im deutschen Teil der Erhebung, der sich auf N = 104 Personen bezieht, interessiert uns in diesem Aufsatz in erster Linie die Reliabilitätsthematik. Diese Thematik umfaßt die folgenden Fragestellungen:

- Welche Berufsmerkmale können im Rahmen eines geeigneten Meßmodells als Indikatoren welcher Konstrukte aufgefaßt werden, bzw. welche latenten Konstrukte werden durch die Wichtigkeitseinstufungen welcher Berufsmerkmale erfaßt?
- Wie rein werden die Konstrukte durch die entsprechenden Merkmale erfaßt? Sind die Items in bezug auf die erfaßten Konstrukte kongenerisch, -äquivalent oder sogar nominell parallel? (Zur Klärung der Begriffe vgl. Abschnitt 2.) Gibt es zwischen den verbal und numerisch erfaßten Merkmalen Reliabilitätsunterschiede?
- Inwieweit läßt sich durch eine Gewichtung der ein bestimmtes Konstrukt erfassenden Items die Konstruktmessung verbessern?

Zwischen der Darbietung der numerischen Skala und der Darbietung der verbalen Skala lag eine längere, die Konzentration der Person voll beanspruchende Aufgabe, bei der für alle möglichen Paare von Merkmalen angegeben werden mußte, welches Merkmal für wichtiger gehalten wird und für wievielfach wichtiger es gehalten wird, sodaß der Erinnerungsfaktor als weitgehend ausgeschaltet gelten konnte. Andererseits lagen beide Aufgaben aber zeitlich so

eng beisammen, daß eine inzwischen eingetretene Änderung der latenten Einstellungsstrukturen nicht zu erwarten war, so daß insgesamt gesehen eine geradezu ideale Test-Retest-Situation vorlag.

Im Zusammenhang mit der numerischen Version der Skala interessieren uns natürlich auch die entsprechenden Reliabilitäten der ALLBUS-Hauptuntersuchung, über die hier zum ersten Mal berichtet wird.

Die konfirmatorischen Faktorenanalysen, die dieser Arbeit zugrundeliegen, wurden ausnahmslos mit dem Programm LISREL V gerechnet (vgl. JÖRESKOG & SÖRBOM, 1981). Dabei folgten wir der in weiten Kreisen üblichen Praxis, Ordinaldaten in Modellen mit multiplen Indikatoren wie Intervalldaten zu behandeln. Zu diesem Problem liegen eine Reihe z. T. kontroverser Stellungnahmen vor (vgl. z. B. BOLLEN & BARB, 1981; HENRY, 1982). Neuere Simulationsstudien kommen jedoch zu dem Ergebnis, daß sich bei der Verwendung von mehr als vier Kategorien die Parameterschätzungen bei Ordinalskalen von den Schätzungen der entsprechenden kontinuierlichen Skalen kaum unterscheiden (vgl. JOHNSON & CREECH, 1983).

2. Reliabilität und konfirmatorische Faktorenanalyse

Nach der klassischen Testtheorie (vgl. z. B. LORD & NOVICK, 1968, Kap. 3) kann der Wert x_i , den ein Proband in einem Test i erhält, in zwei Komponenten zerlegt werden: in einen wahren Wert (true score) τ_i und einen Meßfehler ϵ_i ; in Symbolen:

$$x_i = \tau_i + \epsilon_i \quad (1)$$

Hierbei wird angenommen, daß die wahren Werte und die Fehlerwerte des gleichen Tests sowie die Fehlerwerte verschiedener Tests nicht korreliert sind; d. h. es gilt:

$$\text{Cov}(\tau_i, \epsilon_i) = 0 \quad \text{und} \quad \text{Cov}(\epsilon_i, \epsilon_j) = 0, \quad (2)$$

wobei $\text{Cov}(x,y)$ die Kovarianz zwischen x und y bezeichnen möge. Ferner gilt:

$$E(\epsilon_i) = 0 \quad (3)$$

d. h. der Erwartungswert des Meßfehlers über eine theoretisch unendliche Zahl von Wiederholungen des gleichen Tests am gleichen Probanden ist 0.

Aus diesen Voraussetzungen folgt, daß die Varianz der beobachteten Werte eines Tests x (beobachtete Varianz) gleich der Summe der Varianz der wahren Werte (wahre Varianz) und der Varianz der Fehlerwerte (Fehlervarianz) ist:

$$\sigma_x^2 = \sigma_\tau^2 + \sigma_\epsilon^2 \quad (4)$$

Die Reliabilität eines Tests x , ρ_x , ist dann definiert als Verhältnis der wahren Varianz zur beobachteten Varianz:

$$\rho_x = \frac{\sigma_\tau^2}{\sigma_x^2} \quad (5)$$

Wie man durch Umformung leicht erkennt, ist die Korrelation $\rho_{\tau x}$ zwischen den wahren Werten und den beobachteten Werten

$$\rho_{\tau x} = \frac{\text{Cov}(\tau, x)}{\sigma_\tau \sigma_x} \quad (6)$$

gleich der Quadratwurzel der Reliabilität. Diese Korrelation wird gelegentlich auch als theoretische Validität bezeichnet (vgl. z. B. LORD & NOVICK, 1968:261; BOHRNSTEDT, 1983:8). Aus (4) und (5) folgt, daß

$$\rho_x = 1 - \frac{\sigma_\epsilon^2}{\sigma_x^2} \quad (7)$$

Zwischen dem Reliabilitätsbegriff der klassischen Testtheorie und der konfirmatorischen Faktorenanalyse hat nun JÖRESKOG (1971, 1978; vgl. auch ALWIN & JACKSON, 1980; CARMINES & McIVER, 1981) einen interessanten Zusammenhang hergestellt. Ausgangspunkt ist der Begriff der kongenerischen Messung. Dabei heißt eine Menge von Messungen x_1, x_2, \dots, x_n kongenerisch genau dann, wenn jedes Paar von wahren Werten τ_i, τ_j mit 1.0 korreliert. Ist dies der Fall, so existiert eine Zufallsvariable τ dergestalt, daß gilt:

$$\tau_i = \mu_i + \beta_i \tau \quad (8)$$

Wegen (1) folgt durch Einsetzung

$$x_i = \mu_i + \beta_i \tau + \epsilon_i \quad (9)$$

Unter Voraussetzung der in der klassischen Testtheorie gültigen Annahmen (2) und (3) und unter Voraussetzung der Annahme, daß τ eine unbekannte Metrik besitzt, so daß es sinnvoll erscheint

$$E(\tau) = 0 \quad \text{und} \quad \text{Var}(\tau) = 1.0$$

zu setzen, ergibt sich

$$V(x_i) = \beta_i^2 + \theta_i^2, \quad (10)$$

wobei θ_i^2 die Fehlervarianz bezeichnet.

Damit folgt aus (5), daß unter den genannten Voraussetzungen die Reliabilität von Test i gegeben ist durch

$$\rho_i = \frac{\beta_i^2}{\beta_i^2 + \theta_i^2} \quad (11)$$

Seien nun \underline{x} , $\underline{\mu}$, $\underline{\beta}$ und $\underline{\varepsilon}$ Spaltenvektoren, die jeweils aus den Werten x_i, μ_i, β_i und ε_i bestehen. Dann kann (9) in Vektorform geschrieben werden als

$$x_i = \underline{\mu} + \underline{\beta}\tau + \underline{\varepsilon} \quad (12)$$

Bezeichnet man mit $\underline{\Sigma}$ die Varianz-Kovarianzmatrix von \underline{x} und mit $\underline{\theta}^2$ die Diagonalmatrix der Fehlervarianzen, so ergibt sich

$$\underline{\Sigma} = \underline{\beta}\underline{\beta}' + \underline{\theta}^2 \quad (13)$$

(13) ist aber die Grundgleichung eines Faktormodells mit einem gemeinsamen Faktor. Unter der Voraussetzung, daß \underline{x} eine multivariate Normalverteilung besitzt, lassen sich die Parameter in (13) effizient schätzen. Über den χ^2 -Test der Anpassungsgüte läßt sich das kongenerische Meßmodell statistisch überprüfen.

Als Spezialfälle des Modells kongenerischer Tests erhalten wir das Modell τ -äquivalenter Tests und das Paralleltest-Modell: Tests 1,2, ...,n heißen τ -äquivalent genau dann, wenn gilt

$$\beta_1 = \beta_2 = \dots = \beta_n \quad (14)$$

Sie heißen nominal parallel genau dann, wenn zusätzlich zu (14) gilt, daß

$$\theta_1 = \theta_2 = \dots = \theta_n \quad (15)$$

τ -äquivalente Tests zeichnen sich also durch gleiche wahre Varianzen, aber möglicherweise unterschiedliche Fehlervarianzen aus, während im Falle nomineller Parallelität zusätzlich noch die Gleichheit der Fehlervarianzen besteht. Eine noch strengere Form der Parallelität wäre dann gegeben, wenn wir zusätzlich zu (15) noch annehmen würden, daß

$$\mu_1 = \mu_2 = \dots = \mu_n \quad (16)$$

Die Parameter von durch die Annahmen (14) und (15) eingeschränkten Modellen lassen sich mit Hilfe der Programme COFAMM (JÖRESKOG & SÖRBOM, 1976) oder LISREL schätzen.

Reliabilität gewichteter Testwerte:

Ist eine Menge von Tests bzw. Items kongenerisch, so läßt sich aus ihnen ein gewichteter Score

$$y = \underline{w}'\underline{x} \quad (17)$$

bilden. Hierbei ist \underline{w}' ein $1 \times n$ Vektor von Gewichten. Setzt man (9) in (17) ein, so erhält man

$$y = \underline{w}'\underline{\mu} + (\underline{w}'\underline{\beta}) + \underline{w}'\underline{\varepsilon} \quad (18)$$

Bestimmt man die Varianz von (18) und wendet man die Reliabilitätsdefinition von (5) an, so erhält man, wie JÖRESKOG (1971) zeigt, als Reliabilität von y

$$\rho = \frac{\underline{w}' \underline{\beta} \underline{\beta}' \underline{w}}{\underline{w}' \underline{\beta} \underline{\beta}' \underline{w} + \underline{w}' \underline{\theta}^2 \underline{w}} \quad (19)$$

Man kann ferner zeigen, daß ρ maximal wird, wenn man

$$w_i = \frac{\beta_i}{\theta_i^2} \quad (20)$$

setzt.

Die w_i sind also jene Gewichte, mit denen man unter den gegebenen Umständen in bezug auf das zur Diskussion stehende Konstrukt maximale Reliabilität erreichen kann.

3. Die Ergebnisse der exploratorischen Faktorenanalysen

Um Aussagen über die Itemreliabilitäten zu machen, erscheint es sinnvoll, sich zunächst einen Überblick über die latenten Konstrukte zu verschaffen, die durch die Wichtigkeitseinstufungen der beruflichen Merkmale erfaßt werden. Einen solchen Überblick gewinnen wir mit Hilfe entsprechender exploratorischer Faktorenanalysen. In Tabelle 1 sind die Ergebnisse für die numerische Skala des ALLBUS 1982, die numerische Skala der Methodenstudie und die verbale Skala der Methodenstudie zusammengestellt. Die dort angegebenen Ladungen sind Varimax-rotiert und wurden bestimmt mit Hilfe einer mit SPSS gerechneten Hauptkomponentenanalyse ohne iterierte Kommunalitäten. Die vier höchsten Ladungen auf einem Faktor sind jeweils unterstrichen.

Wie man sieht, ergaben die Analysen für alle drei Skalen eine relativ klare dreidimensionale orthogonale Struktur, die der schon im ALLBUS 1980 gefundenen Struktur weitgehend entspricht (vgl. SCHMIDT, 1982). Die einzelnen Faktoren wurden dort als "Extrinsische Orientierung" (EXT), "Intrinsische Orientierung" (INT) und "Soziale Orientierung" (SOZ) bezeichnet.

Tab. 1: Ergebnisse der exploratorischen Faktorenanalysen

<u>Variablen</u>	ALLBUS 1982 NUMERISCH			SKALENVERGLEICH NUMERISCH			SKALENVERGLEICH VERBAL		
	EXT	INT	SOZ	EXT	INT	SOZ	EXT	INT	SOZ
Sichere Berufsstellung	<u>0.714</u>	0.348	0.114	<u>0.726</u>	-0.033	0.270	<u>0.714</u>	-0.162	<u>0.311</u>
Hohes Einkommen	<u>0.835</u>	0.085	0.070	<u>0.864</u>	0.011	-0.179	<u>0.839</u>	-0.086	-0.221
Gute Aufstiegsmöglichkeiten	<u>0.807</u>	0.249	0.174	<u>0.847</u>	0.211	0.113	<u>0.873</u>	0.010	0.018
Ein Beruf, der anerkannt und geachtet wird	<u>0.697</u>	0.295	0.280	<u>0.694</u>	0.051	0.280	<u>0.666</u>	-0.049	0.271
Ein Beruf, der einem viel Freizeit läßt	0.517	0.352	0.113	0.006	0.442	-0.409	-0.100	0.190	-0.559
Interessante Tätigkeit	0.371	<u>0.771</u>	0.184	-0.073	<u>0.743</u>	0.041	-0.080	<u>0.762</u>	-0.253
Eine Tätigkeit, bei der man selbständig arbeiten kann	0.241	<u>0.831</u>	0.177	0.036	<u>0.821</u>	0.060	-0.040	<u>0.778</u>	-0.242
Aufgaben, die viel Verantwortungsbewußtsein erfordern	0.295	<u>0.669</u>	0.340	0.221	<u>0.708</u>	0.223	0.031	<u>0.669</u>	0.124
Viel Kontakte zu anderen Menschen	0.087	0.355	<u>0.637</u>	0.140	<u>0.537</u>	0.385	-0.086	<u>0.631</u>	0.302
Ein Beruf, bei dem man anderen helfen kann	0.148	0.156	<u>0.877</u>	0.199	0.009	<u>0.840</u>	-0.038	0.381	<u>0.746</u>
Ein Beruf, der für die Gesellschaft nützlich ist	0.207	0.175	<u>0.831</u>	0.342	-0.059	<u>0.738</u>	0.044	-0.102	<u>0.848</u>
Gibt mir das Gefühl, etwas Sinnvolles zu tun	0.176	<u>0.601</u>	<u>0.517</u>	-0.025	0.349	<u>0.707</u>	-0.269	0.570	<u>0.353</u>
Sichere und gesunde Arbeitsbedingungen	0.401	0.574	0.297	0.022	0.218	<u>0.419</u>	0.076	0.076	0.258

Betrachtet man die Durchschnittsmengen der vier höchsten Ladungen über alle drei Skalen, so ergeben sich für die einzelnen Konstrukte die folgenden Kernitems:

1. Extrinsische Orientierung

Sichere Berufsstellung

Hohes Einkommen

Gute Aufstiegsmöglichkeiten

Ein Beruf, der anerkannt und geachtet wird

2. Intrinsische Orientierung

Interessante Tätigkeit

Eine Tätigkeit, bei der man selbständig arbeiten kann

Aufgaben, die viel Verantwortungsbewußtsein erfordern

3. Soziale Orientierung

Ein Beruf, bei dem man anderen helfen kann

Ein Beruf, der für die Gesellschaft nützlich ist

Der Faktor "Extrinsische Orientierung" scheint also im wesentlichen den Karriereaspekt zu kennzeichnen, während der Faktor "Intrinsische Orientierung" im wesentlichen den Aspekt der persönlichen Erfüllung kennzeichnet. Der Faktor "Soziale Orientierung" betrifft das persönliche Engagement für die Gesellschaft.

Neben den über die Analysen hinweg invarianten Kernitems gibt es im Falle der "Intrinsischen Orientierung" und der "Sozialen Orientierung" weitere Items, die diese Faktoren in unterschiedlichem Ausmaß charakterisieren. So gehört im Falle der ALLBUS-Untersuchung das Item "Gibt mir das Gefühl, etwas Sinnvolles zu tun" zu den vier höchsten Ladungen, während dies bei den Skalen der Methodenstudie nicht der Fall ist. Dort gehört dagegen das Item "Viel Kontakt zu anderen Menschen" zu den vier höchsten Ladungen. Ähnliches ergibt sich, wenn wir den sozialen Faktor betrachten, wo das Item "Viel Kontakte zu anderen Menschen" im Falle des ALLBUS zu den vier höchsten Ladungen gehört, nicht aber im Falle der beiden Skalen der Skalenvergleichsstudie. Diese Unterschiede sind möglicherweise auf eine unterschiedliche Besetzung der beiden Stichproben zurückzuführen.

4. Die Reliabilität der Einzelitems und die Bildung optimal gewichteter Skalen

Nachdem wir uns - einer verbreiteten Praxis folgend (vgl. z. B. JÖRESKOG, 1971; BOHRNSTEDT, 1983) - mittels exploratorischer Faktorenanalysen einen Überblick über die den Items zugrundeliegenden latenten Konstrukte verschafft haben, wollen wir nun der Frage nachgehen, wie rein diese Konstrukte durch die einzelnen Items erfaßt werden und wie stark die Reliabilität der Messung eines Konstrukts verbessert werden kann, wenn man die Items, die dieses Konstrukt substantiell charakterisieren, zu einer nach Gl.(20) (Abschnitt 2) optimal gewichteten Skala zusammenfaßt. Dabei wurden für ein Konstrukt nur solche Items in Betracht gezogen, die bei mindestens einer der drei Skalen zu den vier höchsten Ladungen zählen und eine Ladung größer als 0.6 besitzen, so daß unter den einzelnen Konstrukten folgende Items subsummiert werden konnten:

1. Extrinsische Orientierung

Sichere Berufsstellung
Hohes Einkommen
Gute Aufstiegsmöglichkeiten
Ein Beruf, der anerkannt und geachtet wird

2. Intrinsische Orientierung

Interessante Tätigkeit
Tätigkeit, bei der man selbständig arbeiten kann
Aufgaben, die viel Verantwortungsbewußtsein erfordern
Viel Kontakte zu anderen Menschen
Gibt mir das Gefühl, etwas Sinnvolles zu tun

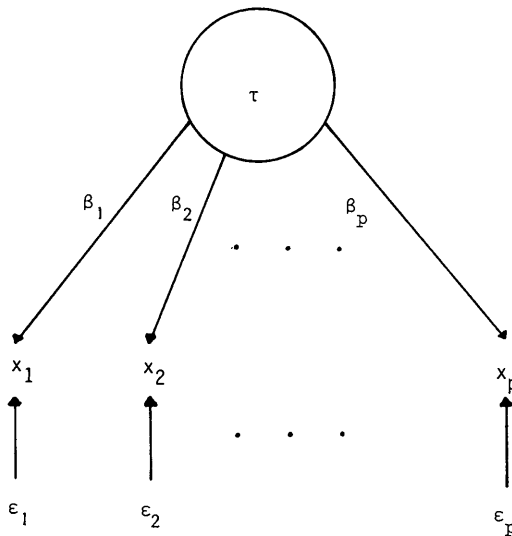
3. Soziale Orientierung

Viel Kontakte zu anderen Menschen
Ein Beruf, bei dem man anderen helfen kann
Ein Beruf, der für die Gesellschaft nützlich ist
Gibt mir das Gefühl, etwas Sinnvolles zu tun

Da für die Bildung einer nach Gl. (20) gewichteten Skala ein kongenerisches Meßmodell vorausgesetzt werden muß, wurde zunächst überprüft, ob diese An-

nahme in bezug auf die von uns betrachteten Konstrukte und die unter ihnen subsummierten Indikatoren akzeptiert werden kann. Dies geschah, indem im Rahmen einer konfirmatorischen Faktorenanalyse das in Abbildung 1 schematisch dargestellte Modell spezifiziert wurde. Die Parameter wurden nach Maximum-Likelihood (ML) geschätzt, wobei als Ausgangsmatrix die Varianz-Kovarianz-Matrix genommen wurde. Die latenten Konstrukte wurden standardisiert.

Abb. 2: Schema eines kongenerischen Meßmodells



Die Ergebnisse der einzelnen Analysen sind in Tabelle 2 wiedergegeben. In der Tabelle sind für jedes Modell die Werte von vier Indizes der Anpassungsgüte zu sehen: χ^2 , χ^2/df , der von JÖRESKOG entwickelte GFI (Goodness-of-Fit-Index) und der RMR (Root Mean Square Residual).

Das χ^2 -Maß ist definiert durch

$$\chi^2 = nF_{\min}, \quad (21)$$

wobei n die Stichprobengröße und F_{\min} das Minimum der Anpassungsfunktion

$$F = \log|\underline{S}| + \text{tr}(\underline{S}\underline{\Sigma}^{-1}) - \log|\underline{S}| - p \quad (22)$$

bezeichnet. In Gl. (22) ist $\underline{\Sigma}$ die theoretische Varianz-Kovarianz-Matrix, \underline{S} die empirische Varianz-Kovarianz-Matrix, $|\underline{A}|$ die Determinante der Matrix \underline{A} , $\text{tr}(\underline{A})$ die Summe der Diagonalelemente von \underline{A} und p die Anzahl der beobachteten Variablen.

Die Anzahl der Freiheitsgrade berechnet sich wie folgt:

$$df = p(p+1) / 2 - t, \quad (23)$$

wobei t die Anzahl der zu schätzenden unabhängigen Parameter ist.

Wie Gl. (21) und (22) zeigen, hängt χ^2 einerseits von der Größe des Unterschieds von theoretisch vorhergesagter und empirisch erhaltener Varianz-Kovarianz-Matrix ab, so daß man versucht wäre, große χ^2 -Werte als Zeichen schlechter Anpassung und kleine χ^2 -Werte als Zeichen guter Anpassung zu interpretieren. Leider erweist sich diese Interpretation aber als vordergründig, da χ^2 andererseits direkt von der Stichprobengröße abhängt. Dies bedeutet, daß die Wahrscheinlichkeit, irgendein Modell abzulehnen, mit wachsender Stichprobengröße zunimmt, wie gut die Übereinstimmung von Modell und Daten auch immer sein mag. Auf der anderen Seite wächst die Wahrscheinlichkeit, ein Modell zu akzeptieren, mit abnehmender Stichprobengröße (vgl. hierzu BENTLER & BONETT, 1980; BOHRNSTEDT & SOBEL, 1983; HOELTER, 1983). Diese Problematik bleibt auch bestehen, wenn man der Daumenregel einiger Autoren (vgl. WHEATON et al., 1977; CARMINES & McIVER, 1981) folgt und das Verhältnis χ^2/df betrachtet.

ZUMA

Tab. 2: Anpassungsindizes für die kongenerischen Modelle

	ALLBUS 1982 NUMERISCH	SKALENVERGLEICH NUMERISCH	SKALENVERGLEICH VERBAL
<u>EXT</u>			
χ^2	5.160 (df=2;P=0.076)	0.860 (df=2;P=0.650)	0.480 (df=2;P=0.786)
χ^2/df	2.580	0.430	0.240
GFI	0.998	0.995	0.997
RMR	0.020	0.041	0.031
<u>INT</u>			
χ^2	78.320 (df=5;P=0.000)	10.290 (df=5;P=0.067)	10.490 (df=5;P=0.062)
χ^2/df	15.660	2.060	2.100
GFI	0.975	0.951	0.850
RMR	0.057	0.117	0.139
<u>SOZ</u>			
χ^2	57.580 (df=2;P=0.000)	6.260 (df=2;P=0.004)	-
χ^2/df	28.790	3.310	-
GFI	0.981	0.968	-
RMR	0.061	0.135	-

Aufgrund der geschilderten Problematik sowie des Sachverhalts, daß χ^2 außerdem sensitiv gegenüber Verletzungen der Normalverteilungsannahme ist, haben wir uns entschlossen, die Anpassungsgüte der Modelle nicht anhand von χ^2 oder χ^2/df , sondern anhand des GFI und des RMR zu beurteilen. Der GFI ist wie folgt definiert:

$$GFI = 1 - \frac{\text{tr}(\hat{\Sigma}^{-1} \underline{S} - \underline{I})^2}{\text{tr}(\hat{\Sigma}^{-1} \underline{S})^2} \quad (24)$$

$\hat{\Sigma}$ bezeichnet die angepaßte Varianz-Kovarianz-Matrix. Der GFI ist ein Maß für den relativen Anteil der Varianzen und Kovarianzen, der durch das Modell gemeinsam erklärt wird. Der GFI sollte zwischen 0 und 1 liegen, obgleich negative Werte theoretisch möglich sind. Er ist unabhängig von der Stichprobengröße und relativ robust gegenüber Abweichungen von der Normalität. Sein Nachteil liegt darin, daß seine Stichprobenverteilung nicht bekannt ist. Der RMR ist ein Maß für den Durchschnitt der Residualvarianzen und -kovarianzen. Er ist definiert durch

$$RMR = \left\{ 2 \sum_{i=1}^k \sum_{j=1}^i (s_{ij} - \delta_{ij})^2 / p(p+1) \right\}^{\frac{1}{2}} \quad (25)$$

Der RMR hat den Nachteil, daß man über ihn Modelle, die an unterschiedlichen Daten überprüft wurden, nicht miteinander vergleichen kann.

Wir glauben - auch angesichts der von uns betrachteten sehr einfachen Modelle - eine konservative Entscheidungsstrategie zu verfolgen, wenn wir bei einem GFI größer 0.950 und einen RMR kleiner 0.100 davon ausgehen, daß das Modell mit den Beobachtungsdaten verträglich ist. Unter Zugrundelegung dieser Entscheidungsregel müssen wir akzeptieren, daß die Annahme eines kongenerischen Meßmodells mit drei Ausnahmen erfüllt ist. Die eine Ausnahme betrifft das Konstrukt der Intrinsischen Orientierung. Dort ist die Anpassung im Falle der verbalen Skala des Skalenvergleichs nach unseren Kriterien nicht als ausreichend anzusehen. Die anderen beiden Ausnahmen betreffen das Konstrukt der Sozialen Orientierung. Dort ist der RMR im Falle der numerischen Skala des Skalenvergleichs nach unserem Maßstab nicht niedrig genug.

Im Falle der verbalen Skala zeigte sich, daß die Evaluation des Modells zu einer "improper solution" in Form negativer Fehlervarianzen führte. Mit derartigen Lösungen ist im Falle kleiner Stichproben gehäuft zu rechnen, wie Simulationsstudien ergaben (vgl. BOOMSMA, 1982). Den RMR von 0.117 im Falle des intrinsischen Konstrukts haben wir gerade noch als zulässig akzeptiert. Insgesamt können wir also sagen, daß nach dem von uns genannten Kriterium im Falle des Konstrukts Extrinsische Orientierung in allen Fällen ein kongenerisches Meßmodell unterstellt werden kann, während die Annahme im Falle der Intrinsischen Orientierung nur für die numerischen Skalen gilt.

Im Falle der Sozialen Orientierung konnten wir das kongenerische Meßmodell nur für den ALLBUS uneingeschränkt akzeptieren.

Wir sahen uns daher berechtigt, in diesen Fällen für die einzelnen Items auf der Basis von Gl. (11) die Reliabilitäten sowie auf der Basis von Gl. (19) die χ^2 -Werte zu berechnen. Die Ergebnisse sind in den Tabellen 3-5 dargestellt, und zwar in Tabelle 3 für das Konstrukt Extrinsische Orientierung, in Tabelle 4 für das Konstrukt Intrinsische Orientierung und in Tabelle 5 für das Konstrukt Soziale Orientierung.

Tab. 3: Reliabilitätsberechnungen für das Konstrukt Extrinsische Orientierung

a. ALLBUS 1982

	β_i	θ_i^2	ρ_i	w_i
Sichere Berufsstellung	1.188	1.075	0.567	1.114
Hohes Einkommen	1.016	1.032	0.502	0.993
Gute Aufstiegs- möglichkeiten	1.417	0.778	0.721	1.821
Ein Beruf, der aner- kannt und geachtet wird	1.223	1.209	0.553	1.012
$\rho = 0.945$				

b. SKALENVERGLEICH NUMERISCH

	β_i	θ_i^2	ρ_i	w_i
Sichere Berufsstellung	1.053	1.600	0.409	0.658
Hohes Einkommen	1.259	1.332	0.543	0.945
Gute Aufstiegs- möglichkeiten	1.276	0.468	0.776	2.726
Ein Beruf, der aner- kannt und geachtet wird	1.093	1.750	0.406	0.625

$\rho = 0.858$

c. SKALENVERGLEICH VERBAL

	β_i	θ_i^2	ρ_i	w_i
Sichere Berufsstellung	0.980	1.508	0.389	0.650
Hohes Einkommen	1.254	1.116	0.585	1.124
Gute Aufstiegs- möglichkeiten	1.229	0.542	0.736	2.268
Ein Beruf, der aner- kannt und geachtet wird	0.880	1.688	0.521	0.315

$\rho = 0.842$

Wie Tabelle 3 zeigt, erreicht im Falle des Konstrukts Extrinsische Orientierung das Item "Gute Aufstiegsmöglichkeiten" sowohl bei der numerisch abgestuften Skala des ALLBUS als auch bei den numerisch bzw. verbal abgestuften Skalen des Skalenvergleichs die höchste Einzelreliabilität, so daß man davon ausgehen kann, daß das zugrundeliegende Konstrukt durch dieses Item am reinsten gemessen wird. Die χ^2 -Werte zeigen, daß durch die verwendete Gewichtung eine substantielle Verbesserung der Konstruktmessung erreicht werden kann. Allerdings weichen die Gewichte teilweise stark voneinander ab. Vergleicht man die numerische Skala des Skalenvergleichs mit der verbalen Skala, so fällt auf, daß im Falle der numerischen Skala die Reliabilität des Items "Ein Beruf, der anerkannt und geachtet wird" erheblich niedriger ausfällt, was aber dank der eingeführten Gewichtung die Gesamtreliabilität aller Items nicht vermindert.

Tab. 4: Reliabilitätsberechnungen für das Konstrukt Intrinsische Orientierung

a. ALLBUS 1982

	β_i	θ_i^2	ρ_i	w_i
Interessante Tätigkeit	1.107	0.643	0.656	1.722
Tätigkeit, bei der man selbständig arbeiten kann	1.154	0.563	0.703	2.050
Aufgaben, die viel Verantwortungsbewußtsein erfordern	1.157	0.866	0.607	1.336
Viel Kontakte zu anderen Menschen	0.827	1.588	0.366	0.521
Gibt mir das Gefühl, etwas Sinnvolles zu tun	0.935	0.952	0.479	0.982

$\rho = 0.878$

b. SKALENVERGLEICH
NUMERISCH

	β_i	θ_i^2	ρ_i	w_i
Interessante Tätigkeit	0.493	0.755	0.243	0.653
Tätigkeit, bei der man selbständig arbeiten kann	0.948	0.481	0.651	1.971
Aufgaben, die viel Verantwortungsbewußtsein erfordern	1.083	0.582	0.668	1.861
Viel Kontakte zu anderen Menschen	0.885	1.962	0.285	1.962
Gibt mir das Gefühl, etwas Sinnvolles zu tun	0.511	1.716	0.132	0.298

$\rho = 0.805$

Tabelle 4 zeigt, daß auch im Falle des Konstrukts Intrinsische Orientierung bei den numerischen Skalen durch die Bildung einer entsprechend gewichteten Skala eine starke Verbesserung der Konstruktmessung erreicht werden kann. Die niedrigsten Reliabilitäten weisen bei beiden Skalen die Items "Viel Kontakte zu anderen Menschen" und "Gibt mir das Gefühl, etwas Sinnvolles zu tun" auf. Im Falle der numerischen Skala der Methodenstudie sind auch die Reliabilitäten für das Item "Interessante Tätigkeit" im Vergleich relativ niedrig. Offensichtlich betont das intrinsische Konstrukt im Falle der Skalenvergleich-Stichprobe mehr das Moment des freien Handlungsspielraums.

Tab. 5: Reliabilitätsberechnungen für das Konstrukt Soziale Orientierung

a. ALLBUS 1982

	β_i	θ_i^2	ρ_i	w_i
Viel Kontakte zu anderen Menschen	0.922	1.412	0.376	0.653
Ein Beruf, bei dem man anderen helfen kann	1.331	0.769	0.697	1.731
Ein Beruf, der für die Gesellschaft nützlich ist	1.273	0.850	0.656	1.498
Gibt mir das Gefühl, etwas Sinnvolles zu tun	0.896	1.022	0.440	0.877
$\rho = 0.848$				

b. SKALENVERGLEICH
NUMERISCH

	β_i	θ_i^2	ρ_i	w_i
Viel Kontakte zu anderen Menschen	0.647	2.327	0.152	0.278
Ein Beruf, bei dem man anderen helfen kann	1.578	0.700	0.781	2.254
Ein Beruf, der für die Gesellschaft nützlich ist	1.223	1.608	0.482	0.761
Gibt mir das Gefühl, etwas Sinnvolles zu tun	0.854	1.248	0.369	0.684
$\rho = 0.840$				

In Tabelle 5 sind die Ergebnisse für den ALLBUS und die numerische Skala des Skalenvergleichs wiedergegeben. Wir haben die Ergebnisse für letztere ebenfalls angegeben, da mancher Leser möglicherweise ein RMR von 0.135 noch für akzeptabel hält. Es zeigt sich, daß das Item "Ein Beruf, bei dem man anderen helfen kann" das Konstrukt am reinsten erfaßt.

Die Reliabilität ist am geringsten bei dem Item "Viel Kontakte zu anderen Menschen" und bei "Gibt mir das Gefühl, etwas Sinnvolles zu tun". Auch hier kann durch eine optimale Gewichtung die Reliabilität der Messung erheblich gesteigert werden.

5. τ -Äquivalenz und Parallelität

Die Ergebnisse der Prüfung auf τ -Äquivalenz und Parallelität sind in Tabelle 6 dargestellt. Die Überprüfung wurde dadurch vorgenommen, daß entsprechend den Definitionen der Gl. (14) und (15) in Abschnitt 2 die Parameter im konfirmatorischen Modell eingeschränkt wurden.

Tab. 6: Überprüfung auf τ -Äquivalenz und Parallelität

a. <u>ALLBUS 1982</u>				
		EXT	INT	SOZ
kongene- nerisch	χ^2	5.160	78.320	57.580
	χ^2/df	2.580	15.660	28.790
	GFI	0.998	0.975	0.981
	RMR	0.020	0.057	0.061
τ -äqui- valent	χ^2	98.670	176.840	221.350
	χ^2/df	19.730	19.650	44.270
	GFI	0.907	0.914	0.863
	RMR	0.223	0.203	0.278
paral- lel	χ^2	113.740	430.060	260.070
	χ^2/df	14.220	33.080	32.510
	GFI	0.910	0.876	0.868
	RMR	0.238	0.177	0.269

ZUMA

b. SKALENVERGLEICH NUMERISCH

		EXT	INT	SOZ
kongene- nerisch	χ^2	0.860	10.290	6.260
	χ^2/df	0.430	2.060	3.130
	GFI	0.995	0.951	0.968
	RMR	0.041	0.117	0.135
τ -äqui- valent	χ^2	3.190	31.540	22.870
	χ^2/df	0.640	3.500	4.570
	GFI	0.967	0.815	0.863
	RMR	0.155	0.232	0.426
parallel	χ^2	20.050	68.070	34.290
	χ^2/df	2.510	5.240	4.290
	GFI	0.920	0.763	0.814
	RMR	0.238	0.392	0.464

c. SKALENVERGLEICH VERBAL

		EXT	INT	SOZ
kongene- nerisch	χ^2	0.480	10.490	-
	χ^2/df	0.240	2.100	-
	GFI	0.997	0.850	-
	RMR	0.031	0.139	-
τ -äqui- valent	χ^2	6.470	13.410	25.860
	χ^2/df	1.290	1.490	5.172
	GFI	0.937	0.942	0.906
	RMR	0.230	0.142	0.364
parallel	χ^2	17.520	42.300	32.990
	χ^2/df	2.190	3.250	4.123
	GFI	0.920	0.852	0.877
	RMR	0.249	0.348	0.346

In Tabelle 6 sind für jedes Modell (kongenerisch, τ -äquivalent und parallel) χ^2 , χ^2/df , GFI und RMR eingetragen. Bleiben wir bei der in Abschnitt 4 eingeführten Entscheidungsregel und akzeptieren nur solche Modelle mit GFI-Werten über 0.950 und RMR kleiner 0.100, gibt es kein Konstrukt in bezug auf das die ihm zugeordneten Items ein τ -äquivalentes oder sogar paralleles Meßmodell bilden würden.

6. Schlußbemerkung

Die hier berichtete Analyse sollte aufzeigen, wie konfirmatorische Deutungen des Reliabilitätsbegriffes in der Praxis nutzbar gemacht werden können (zu weiteren Beispielen s. JÖRESKOG, 1971; ALWIN & JACKSON, 1980). Die berichteten Maximum-Likelihood-Analysen mögen manchem Leser angesichts der kleinen Stichprobengröße als unangemessen erschienen sein. Andererseits zeigen Simulationsstudien (s. BOOMSMA, 1982), daß der Stichprobenbias der Schätzungen bei Stichproben von 100 kaum mehr als 0.02 beträgt. Inhaltlich wurde gezeigt, daß sich im Falle des ALLBUS für alle Konstrukte das kongenerische Modell bewährt hat. Im Falle der Methodenstudie konnte diese Annahme bei der numerischen Skala für das extrinsische und intrinsische Konstrukt bestätigt werden, bei der verbalen Skala nur für das extrinsische Konstrukt. Aufgrund des in Abschnitt 1 geschilderten Aufbaus des Befragungsinstruments müssen wir davon ausgehen, daß diese Unterschiede zwischen numerischer und verbaler Skala im Falle des ALLBUS skalenbedingt sind. Die Annahme der Parallelität und τ -Äquivalenz mußte in allen Fällen zurückgewiesen werden.

Es wurde ferner gezeigt, daß im Falle der kongenerischen Modelle durch die Gewichtung nach JÖRESKOG bei den Wichtigkeitseinstufungen der Berufsmerkmale Reliabilitäten erzielt werden können, die deutlich über den Reliabilitäten der einzelnen Items liegen.

Der vorstehende Bericht wurden von Frank Faulbaum verfaßt, der die Methodenstudie zum ALLBUS 1982 bei ZUMA betreut.

Literatur

- ALWIN, D. F. & JACKSON, D. J. Measurement models for response errors in surveys: Issues and applications. In: K. SCHUESSLER (Ed.). Sociological Methodology 1980. San Francisco: Jossey Bass, 1980.
- BENTLER, P. M. & BONETT, D. G. Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 1980, 588-606.
- BOLLEN, K. & BARB, K. Pearson's R and coarsely categorized measures. American Sociological Review, 46, 1981, 232-239.
- BOOMSMA, A. The robustness of LISREL against small sample sizes in factor analysis models. In: K. G. JÖRESKOG & H. WOLD (Eds.). Systems under indirect observations. Part I. Amsterdam: North-Holland, 1982, 149-173.
- BOHRNSTEDT, G. W. Measurement. In: P. H. ROSSI, J. D. WRIGHT & A. B. ANDERSON (Eds.). Handbook of Survey Research. New York-London: Academic Press, 1983.
- BOHRNSTEDT, G. W. & SOBEL, M. E. The use of null models to evaluate the fit of covariance structure models. Unpublished Manuscript, April, 1983.
- CARMINES, E. G. & McIVER, J. P. Analyzing models with unobserved variables. In: G. W. BOHRNSTEDT & E. F. BOGATTA (Eds.). Social measurement: Current issues. Beverly Hills, California: Sage, 1981, 325-343.
- HENRY, F. Multivariate analysis and ordinal data. American Sociological Review, 47, 1982, 299-307.
- HOELTER, J. W. The analysis of covariance structures. Goodness-of-Fit Indices. Sociological Methods and Research, 11, 1983, 325-344.
- JÖRESKOG, K. G. Statistical analysis of sets of congeneric tests. Psychometrika, 36, 1971, 109-133.
- JÖRESKOG, K. G. Structural analysis of covariance and correlation matrices. Psychometrika, 43, 1978, 443-477.
- JÖRESKOG, K. G. & SÖRBOM, D. COFAMM. User's Guide. National Educational Resources, Inc., 1976.
- JÖRESKOG, K. G. & SÖRBOM, D. LISREL V. User's Guide. National Educational Resources, Inc., 1981.
- JOHNSON, D. R. & CREECH, J. C. Ordinal measures in multiple indicator models. American Sociological Review, 48, 1983, 398-407.
- LORD, F. M. & NOVICK, M. R. Statistical theories of mental test scores. Reading: Addison-Wesley, 1968.
- ROHRMANN, B. Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. Zeitschrift für Sozialpsychologie, 9, 1978, 222-245.
- SCHMIDT, P. Messung von Arbeitsorientierungen: Theoretische Fundierung und Test alternativer Meßmethoden. Analyse und Kritik, 2, 1983 (im Druck).
- WHEATON, B., MUTHEN, B., ALWIN, D. F. & SUMMERS, G. F. Assessing the reliability and stability in panel models. In: D. R. HEISE (Ed.). Sociological Methodology 1977. San Francisco: Jossey-Bass, 1977.